# Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population

R Cui,[1] Y Okada,[2] S G Jang,[3] J L Ku,[3] J G Park,[3] Y Kamatani,[1] N Hosono,[4]
T Tsunoda,[5] V Kumar,[1] C Tanikawa,[1] N Kamatani,[6] R Yamada,[2] M Kubo,[4]
Y Nakamura,[1,7] K Matsuda[1]

## ABSTRACT

**Background and aim** Colorectal cancer (CRC) is a multifactorial disease with both environmental and genetic factors contributing to its development. The incidence of CRC is increasing year by year in Japan. Patients with CRC in advanced stages have a poor prognosis, but detection of CRC at earlier stages can improve clinical outcome. Therefore, identification of epidemiologial factors that influence development of CRC would facilitate the prevention or early detection of disease.

**Methods** To identify loci associated with CRC risk, we performed a genome-wide association study (GWAS) for CRC and sub-analyses by tumour location using 1583 Japanese CRC cases and 1898 controls. Subsequently, we conducted replication analyses using a total of 4809 CRC cases and 2973 controls including 225 Korean subjects with distal colon cancer and 377 controls.

**Results** We identified a novel locus on 6q26-q27 region (rs7758229 in SLC22A3, p=$7.92 \times 10^{-9}$, OR of 1.28) that was significantly associated with distal colon cancer. We also replicated the association between CRC and SNPs on 8q24 (rs6983267 and rs7837328, p=$1.51 \times 10^{-8}$ and $7.44 \times 10^{-8}$, ORs of 1.18 and 1.17, respectively). Moreover, we found cumulative effects of three genetic factors (rs7758229, rs6983267, and rs4939827 in SMAD7) and one environmental factor (alcohol drinking) which appear to increase CRC risk approximately twofold.

**Conclusions** We found a novel susceptible locus in SLC22A3 that contributes to the risk of distal colon cancer in an Asian population. These findings would further extend our understanding of the role of common genetic variants in the aetiology of CRC.

## Significance of this study

### What is already known about this subject?
► Inherited susceptibility was thought to account for ∼35% of all colorectal cancer (CRC).
► Genome-wide association studies (GWASs) among European populations have identified a subset of CRC susceptibility loci.
► Different genetic factors may contribute to the pathogenesis of distal colon cancer and proximal colon cancer.

### What are the new findings?
► A GWAS for CRC was conducted for the first time in an Asian population and identified a novel susceptible locus in SLC22A3 that was significantly associated with distal colon cancer.
► Genetic and environmental factors play a crucial role in the development of CRC in this Asian population.
► Interesting racial diversity between Asian and Caucasian populations in the pathogenesis of colorectal cancer was suggested.

### How might it impact on clinical practice in the foreseeable future?
► Our findings would contribute to the understanding of the cumulative role of genetic and environmental factors in colorectal carcinogenesis in the Asian population and also establish personalised medical treatment.

## INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer death worldwide. The recent development of novel drugs and therapeutics has remarkably improved overall survival, and the 5-year survival rate is around 90% for patients who are diagnosed at stage I.[1] However, the prognosis of patients with CRC in an advanced stage is still disappointing. Hence, identification of epidemiological factors that influence development of the disease would facilitate its prevention or early detection and subsequently provide better prognosis.

Family history is acknowledged to be one of the strong risk factors, and an approximately twofold increased risk for CRC was observed among patients who have a first-degree relative with CRC.[2] Thus, nearly 15% of patients with CRC have a positive family history of disease.[3–5] Although inherited susceptibility was thought to account for ∼35% of all CRC,[6] high-risk germline mutations in APC, DNA mismatch repair genes, MUTYH, SMAD4, BMPR1A, and LKB1 account for <6% of all cases.[7] Therefore, the remaining heritable CRC risk (approximately 30%) would be caused by the combination of common variants with modest effects.

Recent genome-wide association studies (GWASs) among European populations have identified a subset of CRC susceptibility loci in 8q24,[8 9] 8q23.3 (EIF3H),[10] 10p14,[10] 11q23,[11] 15q13,[12] 18q21 (SMAD7),[11] [13] 14q22.2 (BMP4),[14] 16q22.1 (CDH1),[14] 19q13.1(RHPN2),[14] and 20p12.3.[14] However, no GWAS for CRC in an Asian

population has been performed. To explore the variants that predispose to the disease in an Asian population, we conducted a GWAS for CRC and sub-analyses by tumour location.

## METHODS
### Study population
In this study, we used a total of 6167 CRC cases and 4494 control subjects from a Japanese population. CRC patients were categorised into two main groups (colon and rectal cancers) according to the tumour location. For the colon cancer, we further categorised into proximal colon (caecum, ascending colon, and transverse colon) and distal colon (descending colon, and sigmoid colon) cancers. Characteristics of each cohort are shown in table 1. Case samples of the Japanese population were obtained from BioBank Japan (http://biobankjp.org/).[15] Control DNA samples in the screening stage were obtained from healthy volunteers (n=904, 74.4% males, mean age at diagnosis=52.5 years, SD±14.3) as well as from BioBank Japan (n=994, keloidosis, chronic hepatitis B, pulmonary tuberculosis, and drug rash). To increase the power to detect genetic factors related with CRC, we used cases with higher hereditary predisposition in our screening stage. CRC cases who developed the disease at a younger age (≤60 years old) or had at least one first-degree relative with a history of CRC were enriched in our screening stage cohort (62.80% and 35.5%, respectively) compared to the first replication (21.17% and 3.62%, respectively) and the second replication cohorts (21.62% and 8.28%) (supplementary table 1). Genotyping results of 2596 individuals that were registered in BioBank Japan were used as the controls for the first (ischaemic stroke and myocardial infarction) and the second (peripheral artery disease and arrhythmia) replication analyses. We excluded the subjects with cancers or diabetes from control groups. Patients with inflammatory bowel diseases (ulcerative colitis and Crohn's disease) were not included as either cases or controls. Cases and controls (healthy volunteers) for the third replication (n=225 and 377, respectively) were collected at Cancer Research Institute, Seoul National University College of Medicine, Korea. All the participants provided written informed consent.

### SNP genotyping and quality control
Platforms used in each stage are shown in table 1. Analyses using Illumina Beadchip (Illumina, San Diego, California, USA) were conducted at the Center for Genomic Medicine, The Institute of Physical and Chemical Research. In the first and the second replication analyses, genotyping of case samples was conducted by the multiplex polymerase chain reaction-based Invader assay (Third Wave Technologies, Madison, Wisconsin, USA).[16] Korean samples in the third replication analysis were genotyped using the Taqman assay (Applied Biosystems, San Francisco, California, USA). To evaluate the quality of each genotyping methods, we analysed rs7758229 in 96 CRC cases. Consequently, we observed more than 99% concordance between the result from direct sequencing and those from three genotyping systems (invader, Taqman, and Illumina Human610-Quad BeadChip). These findings indicated that different genotyping methods are not likely to cause the inflation of association in our study.

In the screening stage, we genotyped 1595 CRC cases and 1903 control subjects. These samples were genotyped using the Illumina Human610-QuadBeadChip in cases and the Illumina HumanHap550v3 BeadChip in controls. In fact, genotype concordance between these two BeadChips was 99.99% among

182 duplicate samples, indicating a low possibility of genotype error. The samples with a call rate of < 0.98 were excluded from our analysis (12 cases and five controls). Then we applied SNP quality control as follows: call rate ≥0.99 in cases and controls, Hardy−Weinberg p≥$1×10^{-7}$ in controls. Finally, 496 531 common SNPs between Human610-Quad Beadchip and HumanHap550v3 Genotyping BeadChip on autosomal chromosomes passed the quality control criteria. We selected 391 749 SNPs with minor allele frequency (MAF) of ≥0.05 in either case or control samples for further analyses, considering statistical power in the replication analyses.

### Statistical analysis
In the screening stage, the associations between each SNP and CRC were assessed using the Cochran−Armitage trend test. Thirty-six SNPs that exhibited false discovery rate $Q$ value ≤0.2 (p≤$2.5×10^{-5}$) were further analysed using an independent cohort consisting of 3099 CRC cases and 1777 controls. Sub-analyses by tumour location were also performed applying the same criteria (p≤$2.5×10^{-5}$). The significance thresholds were set to be 0.05 in the first, second and third replication study, and $1.27×10^{-7}$ (0.05/391 749) in the meta-analysis. When we take account of the sub-group analyses for multiple testing correction, the genome-wide significance threshold is $2.54×10^{-8}$ (0.05/(391 749×5)). The statistical powers to detect a variant with OR of 1.3 and MAF of 0.2 were estimated to be 62.6%, 99.9% and 94.2% (screening stage, first and second replication) for CRC, and 19.1%, 95.8%, and 77.2% for distal colon cancer, respectively. ORs and CIs were calculated using the major allele as a reference. Since alcohol intake of more than two standard drinks (28 g of pure alcohol per day) was shown to increase the risk of CRC,[17] we classified the subjects into three categories: non-drinkers (0−1 g/day of alcohol), light drinkers (1−28 g/day of alcohol), or heavy drinkers (≥28 g/day of alcohol). Multiple logistic regression analysis was used to assess the contributions of the confounding factors with the R program (version 2.8.1). Age and gender were designated as regulatory factors, and the following explanatory variables were included in the analysis: alcohol consumption status (0=non-drinker, 1=light drinker, 2=heavy drinker), tobacco smoking (0=never smoker, 1=smoker) and rs7758229 genotype (0=GG, 1=TG, 2=TT).

### Imputation analysis
To investigate the region including SNP rs7758229, imputation of ungenotyped SNPs was conducted by MACH v1.0 (see URLs) using first screening GWAS dataset with reference to the release 27 JPT (Japanese in Tokyo, Japan) from the HapMap Project website. Imputed SNPs with an Rsq value <0.3 were excluded. We also genotyped two exonic SNPs (rs668871 and rs2292334) in 48 cases and 48 controls by direct sequencing and found that the results from imputation analysis were completely consistent with those from sequence analysis.

### SNP discovery
We performed re-sequencing of the promoter region and all exons of the SLC22A3 gene using genomic DNA from 48 individuals with distal colon cancer and 48 healthy controls. Primers used for amplification and sequence analyses were indicated in the supplementary table 9.

## RESULTS
To identify common variants that influence CRC risk, we genotyped 1583 Japanese individuals with CRC and 1898

**Table 1** Characteristics of study populations

| Cohort | Source | Platform | Number of samples | Male | Age (mean±SD) | Drinking status | | |
|--------|--------|----------|-------------------|------|---------------|-----------------|---|---|
| | | | | | | Nondrinker | Light drinker | Heavy drinker |
| *Screening stage* | | | | | | | | |
| **Case** | | | | | | | | |
| Colorectal cancer | BioBank Japan | Illumina HumanHap 610k | 1583 | 60.83% | 58.37±10.91 | 53.38% | 25.65% | 20.97% |
| Colon cancer | BioBank Japan | Illumina HumanHap 610k | 1025 | 58.44% | 59.33±11.20 | 55.51% | 23.90% | 20.59% |
| Proximal colon cancer | BioBank Japan | Illumina HumanHap 610k | 410 | 54.15% | 60.79±11.75 | 60.24% | 22.44% | 17.32% |
| Distal colon cancer | BioBank Japan | Illumina HumanHap 610k | 579 | 60.97% | 58.07±10.49 | 52.68% | 24.70% | 22.62% |
| Rectal cancer | BioBank Japan | Illumina HumanHap 610k | 497 | 63.98% | 55.78±9.82 | 50.10% | 28.37% | 21.53% |
| **Control** | Midosuji Rotary Club and BioBank Japan | Illumina HumanHap 550k | 1898 | 63.78% | 52.49±14.38 | 52.50% | 34.37% | 13.13% |
| *First replication* | | | | | | | | |
| **Case** | | | | | | | | |
| Colorectal cancer | BioBank Japan | Invader assay | 3099 | 63.52% | 66.77±8.76 | 59.57% | 24.07% | 16.36% |
| Colon cancer | BioBank Japan | Invader assay | 1618 | 62.16% | 67.64±8.55 | 60.19% | 24.42% | 15.39% |
| Proximal colon cancer | BioBank Japan | Invader assay | 737 | 58.61% | 68.90±8.72 | 63.31% | 23.03% | 13.66% |
| Distal colon cancer | BioBank Japan | Invader assay | 848 | 64.62% | 66.62±8.25 | 57.06% | 25.84% | 17.10% |
| Rectal cancer | BioBank Japan | Invader assay | 869 | 66.55% | 65.85±8.28 | 57.29% | 25.79% | 16.92% |
| **Control** | BioBank Japan | Illumina HumanHap 610k | 1777 | 76.14% | 60.56±10.33 | 57.45% | 27.35% | 15.20% |
| *Second replication* | | | | | | | | |
| **Case** | | | | | | | | |
| Colorectal cancer | BioBank Japan | Invader assay | 1485 | 64.75% | 67.39±9.92 | 61.62% | 20.52% | 17.86% |
| Colon cancer | BioBank Japan | Invader assay | 854 | 63.58% | 68.18±9.63 | 62.11% | 20.91% | 16.98% |
| Proximal colon cancer | BioBank Japan | Invader assay | 355 | 58.31% | 69.01±9.95 | 66.47% | 19.20% | 14.33% |
| Distal colon cancer | BioBank Japan | Invader assay | 489 | 66.67% | 67.52±9.35 | 58.59% | 22.57% | 18.84% |
| Rectal cancer | BioBank Japan | Invader assay | 402 | 67.00% | 66.46±9.62 | 62.56% | 18.85% | 18.59% |
| **Control** | BioBank Japan | Illumina HumanHap 610k | 819 | 74.11% | 68.58±11.01 | 59.95% | 21.86% | 18.19% |
| *Third replication* | | | | | | | | |
| **Case** | | | | | | | | |
| Distal colon cancer | Seoul National Univ | TaqMan Genotyping System | 225 | 65.33% | 58.80±12.90 | | | |
| **Control** | Seoul National Univ | TaqMan Genotyping System | 377 | 66.84% | 56.73±10.90 | | | |

CRC patients were categorised into two main groups (colon and rectal cancers) according to the tumour location. For the colon cancer, we further categorised into proximal colon (caecum, ascending colon, and transverse colon) and distal colon (descending colon, and sigmoid colon) cancers. The control samples from BioBank Japan consisted of the individuals with diseases other than cancer. Third replication samples derived from Korea consisted of distal colon cancer patients and healthy controls.

control individuals (supplementary figure 1A). To overcome the relatively low statistical power in the screening stage, patients under 60 years old who developed cancer or who had a positive family history were preferentially enrolled in the screening stage, as mentioned in the methods section. We also performed sub-analysis by tumour location (colon, proximal colon, distal colon and rectal cancer) to explore common variants that predispose to some subsets of CRC (supplementary figure 1B—E). Application of the Cochran—Armitage trend test to all the tested SNPs indicated that the genomic inflation factor λ was 1.05, 1.04, 1.02, 1.03 and 1.02 for the colorectal, colon, proximal colon, distal colon and rectal cancers, respectively (supplementary figure 2A—E). For CRC, the inflation factor λ was 1.05 in our study. Since the inflation factor is highly dependent on sample size,[18 19] we calculated it by using 1000 cases and 1000 controls. As a result, the inflation factor becomes as small as 1.03, implying a low possibility of false positive associations due to population stratification.

SNPs with a false discovery rate Q value ≤0.2 (p≤$2.5×10^{-5}$) in the screening stage were considered as candidates. Thus, 36 SNPs for CRC, 27 SNPs for colon cancer, 20 SNPs for proximal colon cancer, 18 SNPs for distal colon cancer, and nine SNPs for rectal cancer were selected for further analyses (supplementary tables 2—6). In the first replication study, we genotyped these candidate SNPs using independent cohorts consisting of up to

3099 cases and 1777 controls (table 1). Two SNPs for CRC (table 2) and one SNP for distal colon cancer (table 3) exhibited a p value lower than 0.05. Then we analysed these SNPs in 1485 CRC cases, 489 distal colon cancer cases, and 819 controls, respectively. As a result, all three SNPs showed a p value lower than 0.05 in the second replication study. Consequently, we identified a significant association between CRC and SNPs on 8q24 (rs6983267 and rs7837328, p=$1.51×10^{-8}$ and $7.44×10^{-8}$, respectively; table 2), which have been reported to be associated with CRC in studies of Caucasian subjects.[8] In addition, we identified a novel locus (rs7758229 on 6q26-q27) that was significantly associated with distal colon cancer. Meta-analysis of all stages showed a Mantel—Haenszel p value of $1.07×10^{-8}$ and OR of 1.29 (table 3). Since rs7758229 did not associate with any disease that was used in the control group in our study (supplementary figure 4), the case-mix cohort is not likely to affect the association between rs7758229 and distal colon cancer.

For this novel locus, we conducted a replication study using 225 Korean subjects with distal colon cancer and 377 controls. Although the association was not significant, we observed a similar trend in the samples from the Korean subjects (p=0.286 with OR of 1.16) and Mantel—Haenszel p value for independence had improved from $1.07×10^{-8}$ to $7.92×10^{-9}$ (OR of 1.28, $P_{heterogeneity}$=0.20; table 3) when we conducted a meta-analysis of the Japanese and Korean study with a fixed-effects model.

**Table 2** The combined results of screening stage, first replication and second replication analyses for colorectal cancer

| SNP | Allele (1/2) | Chr/gene | Stage | Case | | | | Control | | | | OR (95% CI)† | p Value* |
| | | | | 11 | 12 | 22 | MAF | 11 | 12 | 22 | MAF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs6983267 | (T/G) | 8q24/Not in gene | Screening stage | 593 | 757 | 233 | 38.63 | 826 | 869 | 203 | 33.59 | 1.25 (1.13 to 1.37) | 1.08E−05 |
| | | | First replication | 1214 | 1476 | 408 | 36.99 | 764 | 782 | 231 | 35.00 | 1.09 (1.01 to 1.19) | 4.91E−02 |
| | | | Second replication | 540 | 701 | 239 | 39.83 | 347 | 378 | 94 | 34.55 | 1.26 (1.11 to 1.42) | 4.24E−04 |
| | | | Combination | 2347 | 2934 | 880 | 38.09 | 1937 | 2029 | 528 | 34.32 | 1.18 (1.11 to 1.25) | 1.51E−08 |
| rs7837328 | (A/G) | 8q24/Not in gene | Screening stage | 209 | 750 | 624 | 36.89 | 189 | 842 | 867 | 32.14 | 1.23 (1.12 to 1.36) | 2.78E−05 |
| | | | First replication | 366 | 1448 | 1283 | 35.20 | 209 | 760 | 808 | 33.15 | 1.10 (1.00 to 1.20) | 4.01E−02 |
| | | | Second replication | 215 | 688 | 580 | 37.69 | 86 | 368 | 365 | 32.97 | 1.23 (1.08 to 1.40) | 1.42E−03 |
| | | | Combination | 790 | 2886 | 2487 | 36.23 | 484 | 1970 | 2040 | 32.69 | 1.17 (1.10 to 1.24) | 7.44E−08 |

*p Values derived from Cochran—Armitage trend tests.
†OR of minor allele from 2×2 allele frequency table.
MAF, minor allele frequency

Interestingly, rs7758229 exhibited a much stronger effect on the risk of distal colon cancer among younger populations and the patients with a family history of the disease (supplementary figure 5). Taken together, this association appears to be true.

SNP rs7758229 is located within an intron 5 of *SLC22A3* (the solute carrier family 22, member 3), one of the organic cation transporter genes. Organic cation transporters are critical for the elimination of some drugs and environmental toxins. This SNP is located within a recombination hot spot between two linkage disequilibrium blocks spanning an approximately 350-kb region on chromosome 6q26-q27 (figure 1). To further investigate this candidate region, we conducted imputation analysis using a screening stage genome-wide dataset. As a result, many SNPs within the *SLC22A3* gene locus indicated strong association (supplementary figure 3), suggesting a possible role of SLC22A3 in the pathogenesis of distal colon cancer. To identify the causative variant(s) that might alter the function or expression level of SLC22A3, all exons and the promoter region of *SLC22A3* gene were re-sequenced using genomic DNA from 48 cases and 48 healthy controls. As a result, we identified three novel SNPs consisting of one non-synonymous SNP (novel v1, Serine106-Glycine) in exon 1 and two synonymous SNPs (novel v2, novel v3) in exon 3 and 8 (supplementary table 7). In addition to these three novel SNPs, 19 tag SNPs were genotyped using 1916 distal colon cancer cases (screening stage, the first and the second replication studies) and 1818 controls (screening stage). Although three SNPs in intron 1 (rs884742) and intron 5 (rs3123636 and rs3106164) exhibited suggestive associations with distal colon cancer; no variants indicated stronger association than marker SNP rs7758229 (supplementary table 7).

The cumulative epidemiological evidence revealed that various environmental factors such as alcohol drinking, tobacco smoking, physical activity and diet would affect the risk of CRC.[17][20] Since histories of alcohol drinking and tobacco smoking were available for most of case and control subjects, we performed multiple logistic regression analysis using alcohol drinking and smoking as covariates. As shown in table 4, rs7758229 and alcohol consumption are independent risk factors for distal colon cancer (p=$5.61 \times 10^{-9}$, OR of 1.31 and p=$1.53 \times 10^{-6}$, OR of 1.21, respectively) after adjustment of age and sex.

To assess the impact of SNP rs7758229 on CRC predisposition, we genotyped this variant using all CRC cases. As a result, this variant indicated suggestive association with colon cancer (p=$7.40 \times 10^{-7}$ with OR of 1.21; supplementary table 10) and CRC (p=$1.31 \times 10^{-5}$ with OR of 1.16; supplementary table 11), respectively.

We also analysed CRC loci that were reported in Caucasian GWASs using our screening stage cohort. Among the nine SNPs evaluated, SNP rs6983267 (p=$1.08 \times 10^{-5}$, OR of 1.25) and rs4939827 (p=$9.54 \times 10^{-5}$, OR of 1.25; *SMAD7*) indicated strong association. In addition, SNP rs10795668 showed moderate association (p=$2.47 \times 10^{-2}$, OR of 0.90), while the other six SNPs did not associate with CRC. Similar results were observed for all the nine SNPs when we used only 904 healthy subjects as control, and MAFs of all the nine SNPs in screening stage control samples were almost the same as those in the healthy control samples (supplementary table 8). These results suggest that these three variants are common CRC susceptible loci between Caucasian and Asian populations.

Then we assessed the combined impact of three genetic factors (rs6983267;8q24, rs4939827;*SMAD7*, and rs7758229;*SLC22A3*) and alcohol consumption on the risk of CRC. We assigned a score of 0, 1 or 2 for non-, light, and heavy drinkers, as well as a score of 1 for each risk allele of each SNP. Since patients with a score of 2 were most common in control subjects (30.6%), we used these subjects as a reference (figure 2A). As a result, individuals with a score of 5 or above have an

**Table 3** Results of replication studies and meta-analyses for distal colon cancer

| SNP | Allele (1/2) | Gene | Stage | Case | | | Control | | | OR (95% CI)* | p Value† | p Heterogeneity‡ |
| | | | | 11 | 12 | 22 | 11 | 12 | 22 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs7758229 | (T/G) | *SLC22A3* | Screening stage | 53 | 232 | 292 | 97 | 641 | 1156 | 1.46 (1.26 to 1.70) | 5.58E−07 | |
| | | | First replication | 62 | 313 | 472 | 96 | 603 | 1077 | 1.21 (1.05 to 1.38) | 7.13E−03 | |
| | | | Second replication | 30 | 191 | 268 | 33 | 295 | 490 | 1.22 (1.01 to 1.47) | 3.22E−02 | |
| | | | Meta-analysis (Japanese)§ | | | | | | | 1.29 (1.18 to 1.41) | 1.07E−08 | 0.13 |
| | | | Third replication | 9 | 85 | 130 | 15 | 124 | 238 | 1.16 (0.88 to 1.54) | 2.86E−01 | |
| | | | Meta-analysis (Asian)§ | | | | | | | 1.28 (1.18 to 1.39) | 7.92E−09 | 0.20 |

*OR of minor allele from 2×2 allele frequency table.
†p Values derived from Cochran—Armitage trend tests.
‡Results of Breslow—Day tests.
§Meta-analyses of Japanese and Asian populations, respectively. OR and p value for independence test were calculated by the Mantel—Haenzel method.
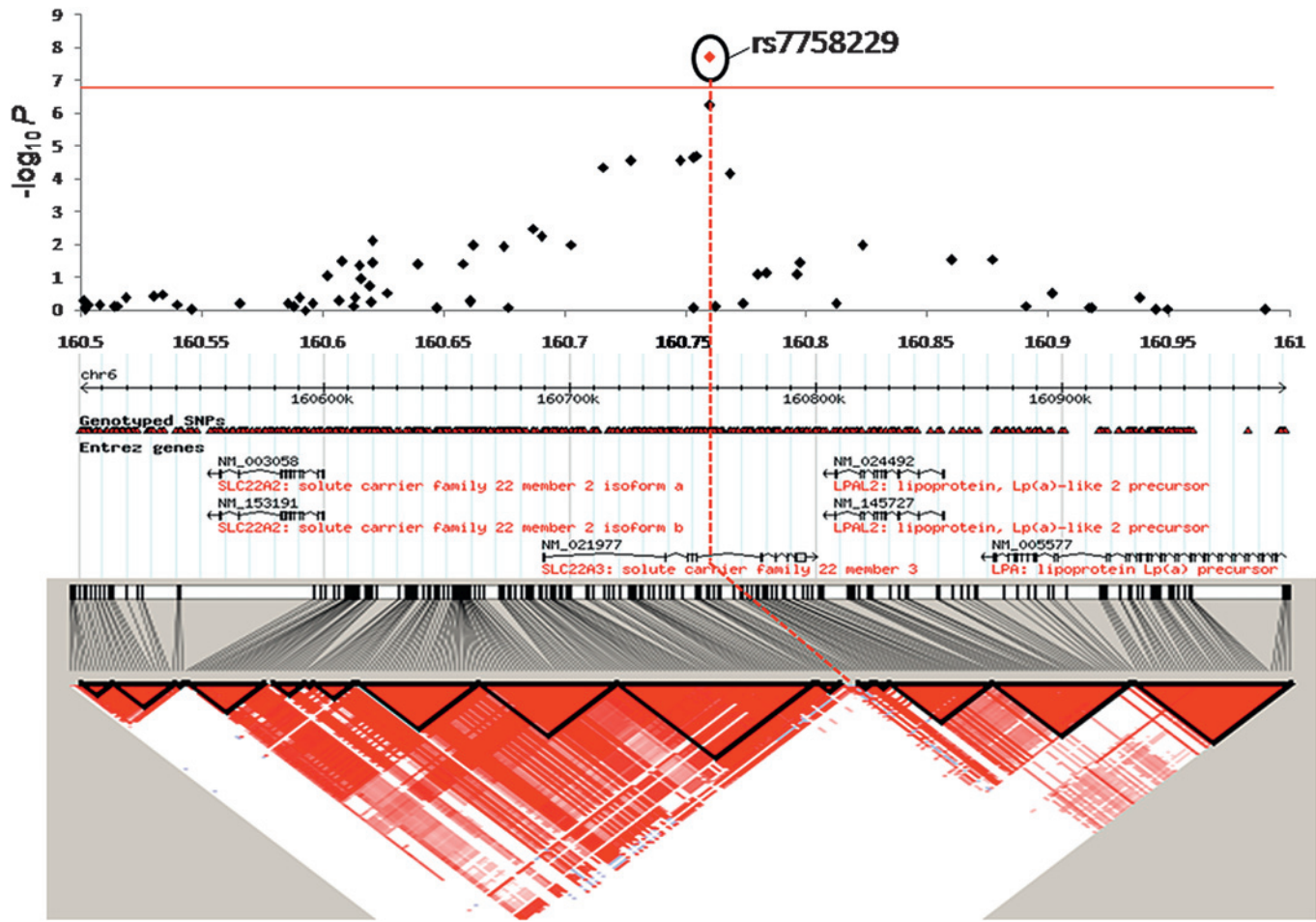
**Figure 1** Case–control association results, genomic structure, and linkage disequilibrium map of chromosomal region of 6q26-q27. The top panel shows single-marker association results calculated by Cochran–Armitage trend tests. Results from the screening stage and combined analysis (screening stage, first and second replication studies) are indicated by black and red dots, respectively. The bottom panel shows the pairwise linkage disequilibrium for SNPs in HapMap JPT. The red line indicates the significance threshold after Bonferroni's correction for multiple testing ($p=1.27\times10^{-7}$).

approximately twofold higher risk of developing CRC compared with individuals with a scores of 2 (figure 2B). These results indicate that genetic and environmental factors play a crucial role in the development of CRC.

## DISCUSSION

CRC that arises proximal or distal to the splenic flexure exhibits differences in incidence according to age, gender and ethnicity. For example, distal colon cancers predominantly occur in white males, while proximal colon cancers are frequent among older African–Americans females.[21][22] The hereditary familial forms of CRC, familial adenomatous polyposis (FAP) and hereditary non-polyposis CRC (HNPCC) also exhibit markedly different clinical features.[23] Nearly 100% of individuals with FAP will develop CRC in the distal colon.[24] In contrast, approximately 70% of large bowel tumours in individuals with HNPCC arise in the proximal colon.[25] In addition, mutations in *TP53* are approximately 1.5- to 3-fold more frequent in distal colon cancer compare with proximal colon cancer.[26] Recently, the incidence of colon cancer in Japan has been increasing.[27] In addition, CRC occurs more frequently in the proximal colon and less frequently in the rectum among the Japanese–American population in Hawaii compared to native Japanese in Japan.[28] These facts suggest that different genetic and environmental factors contribute to the pathogenesis of rectal, distal and proximal colon cancer, respectively.

To date, many studies have shown the associations between various polymorphisms and the risk of CRC, but few have analysed cancer risk by tumour location. In this study, we have identified a novel locus on 6q26-q27, tagged by rs7758229 in *SLC22A3* which was significantly associated with distal colon cancer in Asians. The 6q26-q27 region contains four genes, including *SLC22A2*, *SLC22A3*, *LPAL2* and *LPA*. Imputation analysis of 6q26-q27 region indicated that SNPs around *SLC22A3* revealed strong associations. Interestingly, SNP rs9364554 in intron 5 of *SLC22A3* was shown to associate with prostate cancer in Caucasian populations.[29] Thus *SLC22A3* is likely to be associated with multiple cancers.

*SLC22A3* is a member of the organic cation transporter family that is highly expressed in liver, kidney, intestine and brain.[30]

**Table 4** Logistic regression analysis

| | OR (95% CI) | p Value* |
|---|---|---|
| rs7758229 | 1.31 (1.20 to 1.43) | 5.61E-09† |
| Alcohol consumption | 1.21 (1.12 to 1.31) | 1.53E-06† |
| Smoking | 0.92 (0.80 to 1.05) | 2.10E-01 |

ORs were adjusted for age and sex.
*p Values derived from the Wald test.
†The p value of less than 0.05 was considered statistically significant.
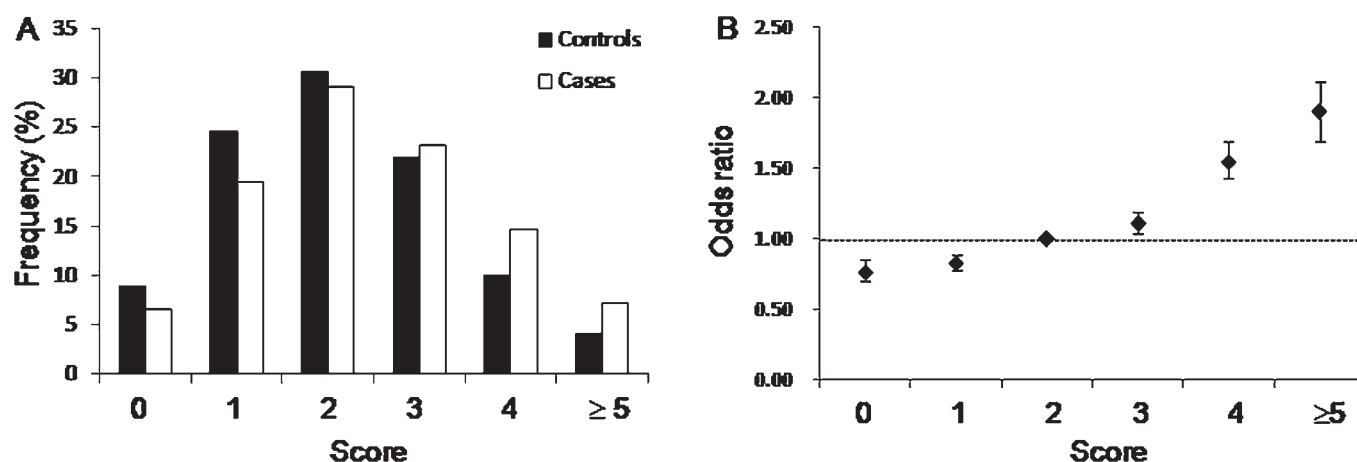
**Figure 2** Impact of genetic and environmental factors on CRC susceptibility. Three genetic factors (rs6983267, rs4939827, and rs7758229) and alcohol consumption were used in the scoring system. We assigned a score of 0, 1, or 2 for non-, light, and heavy drinker, as well as a score of 1 for each risk allele. (A) Distribution of scores in CRC cases (white bars) and controls (black bars). (B) The OR and 95% CIs were indicated relative to the subjects with a score of 2. Horizontal line marks the null value (OR=1). The distribution of scores follows a normal distribution in both cases and controls, with a shift towards higher scores in cases.

This family members play a critical role for the transport of cationic drugs, toxins, and endogenous metabolites.[31] SLC22A3 is expressed in many cancer cell lines, such as colorectal and kidney cancers, and its expression level is correlated with the sensitivity to chemotherapeutic agents.[32 33] Since several toxins or endogenous metabolites such as lipopolysaccharide[34] and linoleic acid metabolite[35] were shown to induce tumour formation, SLC22A3 might be involved in colorectal tumorigenesis through the clearance of some carcinogen.

Numerous studies have indicated that alcohol drinking could be positively associated with CRC risk. In addition, alcohol intake increases the risk of distal colon cancer[17] and/or rectal cancer[36 37] than proximal colon cancer. Furthermore, the effect of alcohol drinking is stronger among Asian populations because of their relatively high prevalence of a slow-metabolising aldehyde dehydrogenase variant.[15 38] Similarly, our study validated that alcohol consumption was strongly associated with distal colon cancer in the Japanese population.

In summary, we have identified a novel susceptible locus in *SLC22A3* that contributes to a risk of distal colon cancer. The incidence of CRC was increasing with the Westernisation of lifestyle and dietary habit in Japan.[39] However, only three of nine CRC susceptibility loci from studies of Caucasians exhibited a p value of less than 0.05 in our study. These results suggest an interesting racial diversity between Asians and Caucasians in the colorectal pathogenesis. Although further functional studies are essential, our findings extend the understanding of the cumulative role of genetic and environmental factors in the colorectal carcinogenesis in the Asian population.

## URLs for items used in the imputation analysis

R statistical environment, http://cran.r-project.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; MACH 1.0, http://www.sph.umich.edu/csg/abecasis/mach/.

## REFERENCES

1. **Jemal A,** Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin* 2008;**58**:71—96.
2. **Johns LE,** Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;**96**:2992—3003.
3. **Carstensen B,** Soll-Johanning H, Villadsen E, et al. Familial aggregation of colorectal cancer in the general population. *Int J Cancer* 1996;**68**:428—35.
4. **Fuchs CS,** Giovannucci EL, Colditz GA, et al. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* 1994;**331**:1669—74.
5. **Slattery ML,** Kerber RA. Family history of cancer and colon cancer risk: the Utah Population Database. *J Natl Cancer Inst* 1994;**86**:1618—26.
6. **Lichtenstein P,** Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;**343**:78—85.
7. **Aaltonen L,** Johns L, Jarvinen H, et al. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res* 2007;**13**:356—61.
8. **Tomlinson I,** Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007;**39**:984—8.
9. **Zanke BW,** Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;**39**:989—94.
10. **Tomlinson IP,** Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008;**40**:623—30.
11. **Tenesa A,** Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 2008;**40**:631—7.
12. **Jaeger E,** Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 2008;**40**:26—8.
13. **Broderick P,** Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 2007;**39**:1315—17.
14. **Houlston RS,** Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008;**40**:1426—35.

15. **Cui R,** Kamatani Y, Takahashi A, *et al*. Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* 2009;**137**:1768—75.

16. **Ohnishi Y,** Tanaka T, Ozaki K, *et al*. A high-throughput SNP typing system for genome-wide association studies. *J Hum Genet* 2001;**46**:471—7.

17. **Thygesen LC,** Wu K, Gronbaek M, *et al*. Alcohol intake and colorectal cancer: a comparison of approaches for including repeated measures of alcohol consumption. *Epidemiology* 2008;**19**:258—64.

18. **Yamaguchi-Kabata Y,** Nakazono K, Takahashi A, *et al*. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet* 2008;**83**:445—56.

19. **Freedman ML,** Reich D, Penney KL, *et al*. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;**36**:388—93.

20. **Cleary SP,** Cotterchio M, Shi E, *et al*. Cigarette smoking, genetic variants in carcinogen-metabolizing enzymes, and colorectal cancer risk. *Am J Epidemiol* 2010;**172**:1000—14.

21. **Gonzalez EC,** Roetzheim RG, Ferrante JM, *et al*. Predictors of proximal vs. distal colorectal cancers. *Dis Colon Rectum* 2001;**44**:251—8.

22. **Kune S,** Kune GA, Watson L. The Melbourne colorectal cancer study: incidence findings by age, sex, site, migrants and religion. *Int J Epidemiol* 1986;**15**:483—93.

23. **Bufill JA.** Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* 1990;**113**:779—88.

24. **Lynch HT,** Smyrk TC. Classification of familial adenomatous polyposis: a diagnostic nightmare. *Am J Hum Genet* 1998;**62**:1288—9.

25. **Lynch HT,** Watson P, Lanspa SJ, *et al*. Natural history of colorectal cancer in hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). *Dis Colon Rectum* 1988;**31**:439—44.

26. **Hamelin R,** Laurent-Puig P, Olschwang S, *et al*. Association of p53 mutations with short survival in colorectal cancer. *Gastroenterology* 1994;**106**:42—8.

27. **Minami Y,** Nishino Y, Tsubono Y, *et al*. Increase of colon and rectal cancer incidence rates in Japan: trends in incidence rates in Miyagi Prefecture, 1959-1997. *J Epidemiol* 2006;**16**:240—8.

28. **Sakamoto K,** Machi J, Prygrocki M, *et al*. Comparison of characteristics and survival of colorectal cancer between Japanese-Americans in Hawaii and native Japanese in Japan. *Dis Colon Rectum* 2006;**49**:50—7.

29. **Eeles RA,** Kote-Jarai Z, Giles GG, *et al*. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008;**40**:316—21.

30. **Kekuda R,** Prasad PD, Wu X, *et al*. Cloning and functional characterization of a potential-sensitive, polyspecific organic cation transporter (OCT3) most abundantly expressed in placenta. *J Biol Chem* 1998;**273**:15971—9.

31. **Inui KI,** Masuda S, Saito H. Cellular and molecular aspects of drug transport in the kidney. *Kidney Int* 2000;**58**:944—58.

32. **Shnitsar V,** Eckardt R, Gupta S, *et al*. Expression of human organic cation transporter 3 in kidney carcinoma cell lines increases chemosensitivity to melphalan, irinotecan, and vincristine. *Cancer Res* 2009;**69**:1494—501.

33. **Yokoo S,** Masuda S, Yonezawa A, *et al*. Significance of organic cation transporter 3 (SLC22A3) expression for the cytotoxic effect of oxaliplatin in colorectal cancer. *Drug Metab Dispos* 2008;**36**:2299—306.

34. **Reisser D,** Pance A, Jeannin JF. Mechanisms of the antitumoral effect of lipid A. *Bioessays* 2002;**24**:284—9.

35. **Yasuda M,** Nishizawa T, Ohigashi H, *et al*. Linoleic acid metabolite suppresses skin inflammation and tumor promotion in mice: possible roles of programmed cell death 4 induction. *Carcinogenesis* 2009;**30**:1209—16.

36. **Poynter JN,** Haile RW, Siegmund KD, *et al*. Associations between smoking, alcohol consumption, and colorectal cancer, overall and by tumor microsatellite instability status. *Cancer Epidemiol Biomarkers Prev* 2009;**18**:2745—50.

37. **Bongaerts BW,** van den Brandt PA, Goldbohm RA, *et al*. Alcohol consumption, type of alcoholic beverage and risk of colorectal cancer at specific subsites. *Int J Cancer* 2008;**123**:2411—17.

38. **Takeshita T,** Morimoto K, Mao X, *et al*. Characterization of the three genotypes of low Km aldehyde dehydrogenase in a Japanese population. *Hum Genet* 1994;**94**:217—23.

39. **Research Group for Population-based Cancer Registration in Japan.** Cancer incidence and incidence rates in Japan in 1997: estimates based on data from 12 population-based cancer registries. *Jpn J Clin Oncol* 2002;**32**:318—22.

# Editor's quiz: GI snapshot

## ANSWERS

*From the questions on page 758*

$T_2$-weighted MRI of the brain demonstrates hazy, symmetrical, bilateral, high signal involving the white matter of the frontal and parietal lobes. This unusual combination of a diffuse leucoencephalopathy with severe gastrointestinal dysmotility,[1][2] cachexia, external ophthalmoplegia and peripheral neuropathy is highly suggestive of the autosomal recessive disorder, mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). This rare disease arises from loss-of-function mutations in the nuclear thymidine phosphorylase (TP) gene, which leads to secondary depletion and deletions in mitochondrial DNA.[3] Gastrointestinal disturbance pre-dates neurological symptoms in >50% of patients and leucoencephalopathy seems to be evident in all patients with confirmed TP deficiency.[4]

Diagnosis was confirmed by demonstrating a marked increase in the normally undetectable purines, thymidine and deoxyuridine, in both the plasma (13 and 17 µmol/l, respectively) and urine (0.2 and 0.3 mmol/l, respectively) and very low white cell TP activity (2 nmol/h/mg protein (normal range: 336—1341)). Sequencing of the TP gene identified compound heterozygosity for two novel mutations (TYMP:c.1167T→C and TYMP:c.1198_1203delGTGCTG). These mutations would lead to an amino acid substitution (p.V389G) and amino acid deletions (p.V400_L401del), respectively, in TP. His parents and younger sibling are phenotypically normal.

Management of MNGIE includes long-term parenteral nutrition and supportive care for the peripheral neuropathy. Though coenzyme Q10 is sometimes prescribed for mitochondrial disorders there is no clear supporting evidence for its routine use in MNGIE.[5] Novel treatments being investigated for MNGIE include dialysis,[6] allogenic stem cell transplantation[7] and administration of carrier erythrocyte-entrapped TP.[8]

## REFERENCES

1. **Blondon H,** Polivka M, Joly F, *et al*. Digestive smooth muscle mitochondrial myopathy in patients with mitochondrial-neuro-gastro-intestinal encephalomyopathy (MNGIE): report of 3 cases and review of the literature. *Gastroentérol Clin Biol* 2005;**29**:773—8.

2. **Mueller LA,** Camilleri M, Emslie-Smith AM. Mitochondrial neurogastrointestinal encephalomyopathy: manometric and diagnostic features. *Gastroenterology* 1999;**116**:959—63.

3. **Nishino I,** Spinazzola A, Hirano M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science* 1999;**283**:689—92.

4. **Hirano M,** Nishigaki Y, Martí R. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): a disease of two genomes. *Neurologist* 2004;**10**:8—17.

5. **Chinnery PF,** Majamaa K, Turnbull D, *et al*. Treatment for mitochondrial disorders. *Cochrane Database Syst Rev* 2006;(1):CD004426.

6. **Yavuz H,** Ozel A, Christensen M, *et al*. Treatment of mitochondrial neurogastrointestinal encephalomyopathy with dialysis. *Arch Neurol* 2007;**64**:435—8.

7. **Hirano M,** Marti R, Casali C, *et al*. Allogeneic stem cell transplantation corrects biochemical derangements in MNGIE. *Neurology* 2006;**67**:1458—60.

8. **Moran NF,** Bain MD, Muqit M, *et al*. Carrier erythrocyte entrapped thymidine phosphorylase therapy for MNGIE. *Neurology* 2008;**71**:686—8.